(51) **International Patent Classification:**
G06T 7/00 (2006.01)

(21) **International Application Number:**
PCT/EP2006/008229

(22) **International Filing Date:** 21 August 2006 (21.08.2006)

(25) **Filing Language:** English

(26) **Publication Language:** English

(30) **Priority Data:**
60/773,714   14 February 2006 (14.02.2006)   US

(71) **Applicant** *(for all designated States except US)*: **FOTO-NATION VISION LIMITED** [IE/IE]; Galway Business Park, Dangen, Galway (IE).

(72) **Inventors; and**

(75) **Inventors/Applicants** *(for US only)*: **CORCORAN, Peter** [IE/IE]; Cregg, Claregalway, Galway (IE). **CAPATA, Adrian** [RO/IE]; 35 Ardilaun Road, Newcastle, Galway (IE). **STEINBERG, Eran** [IL/US]; 137 Granville Way, San Franciso, CA 94127 (US). **BIGIOI, Petronel** [RO/IE]; 57 Sceilg Ard, Headford Road, County Galway (IE).
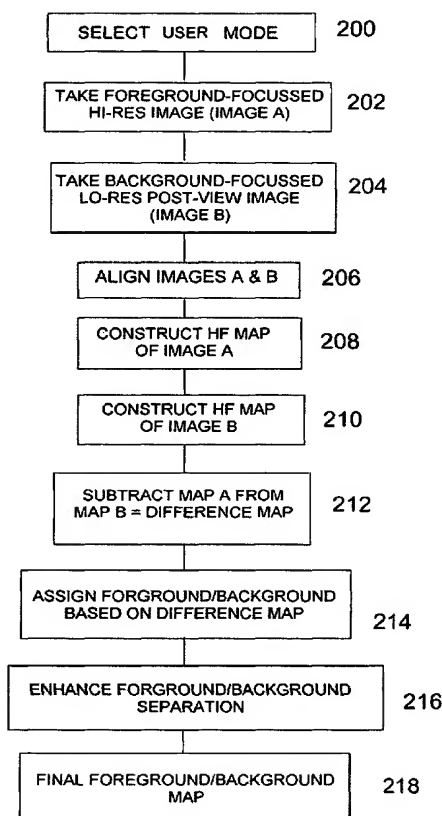
(74) **Agents: CONOR, Boyce** et al.; F.R. Kelly & Co., 27 Clyde Road, Ballsbridge, Dublin 4 (IE).

(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(54) **Title:** FOREGROUND/BACKGROUND SEGMENTATION IN DIGITAL IMAGES

(57) **Abstract:** An implementation efficient method of distinguishing between foreground and background regions of a digital image of a scene comprises capturing two images of nominally the same scene and storing the captured images in DCT-coded format, the first image being taken with the foreground more in focus than the background and the second image being taken with the background more in focus than the foreground. Regions of the first image are assigned as foreground or background according to whether the sum of selected higher order DCT coefficients decreases or increases for the equivalent regions of the second image.

WO 2007/093199 A2 ||||| |||||| || |||| |||| |||| ||||| ||| | ||| |||| |||| ||| ||||| ||| ||| ||||

Foreground/Background Segmentation in Digital Images

This invention related to a method of distinguishing between foreground and background regions of a digital image, known as foreground/background segmentation.

6   BACKGROUND

For some applications the ability to provide foreground/background separation in an image is useful. In PCT Application No. PCT/EP2006/005109 separation based on an analysis of a flash and non-flash version of an image is discussed. However, there are situations where flash and non-flash versions of an image may not provide sufficient discrimination, e.g. in

12   bright sunlight.

Depth from de-focus is a well-known image processing technique which creates a depth map from two or more images with different focal lengths. A summary of this technique can be found at:

http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/FAVARO1/dfdtutorial.html.

18   Favaro is based on a statistical analysis of radiance of two or more images - each out of focus - to determine depth of features in an image. Favaro is based on knowing that blurring of a pixel corresponds with a given Gaussian convolution kernel and so applying an inverse convolution indicates the extent of defocus of a pixel and this in turn can be used to construct a depth map. Favaro requires a dedicated approach to depth calculation once images have been acquired in that a separate radiance map must be created for each image used in depth

24   calculations. This represents a substantial additional processing overhead compared to the existing image acquisition process.

US 2003/0052991, Hewlett-Packard, discloses for each of a series of images taken at different focus distances, building a contrast map for each pixel based on a product of the difference in pixel brightness surrounding a pixel. The greater the product of brightness

30   differences, the more likely a pixel is considered to be in focus. The image with the greatest contrast levels for a pixel is taken to indicate the distance of the pixel from the viewfinder. This enables the camera to build a depth map for a scene. The camera application then implements a simulated fill flash based on the distance information. Here, the contrast map

2

needs to be built especially and again represents a substantial additional processing overhead over the existing image acquisition process.

US 2004/0076335, Epson, describes a method for low depth of field image segmentation. Epson is based on knowing that sharply focussed regions contain high frequency

6   components. US 2003/0219172, Philips, discloses calculating the sharpness of a single image according to the Kurtosis (shape of distribution) of its Discrete Cosine Transform (DCT) coefficients. US 2004/0120598, Xiao-Fan Feng, also discloses using the DCT blocks of a single image to detect blur within the image. Each of Epson, Philips and Feng is based on analysis of a single image and cannot reliably distinguish between foreground and background regions of an image.

12

Other prior art includes US 2003/0091225 which describes creating a depth map from two "stereo" images.

It is an object of the invention to provide an improved method of distinguishing between foreground and background regions of a digital image.

18

DESCRIPTION OF THE INVENTION

According to a first aspect of the present invention there is provided a method of distinguishing between foreground and background regions of a digital image of a scene, the method comprising capturing first and second images of nominally the same scene and

24   storing the captured images in DCT-coded format, the first image being taken with the foreground more in focus than the background and the second image being taken with the background more in focus than the foreground, and assigning regions of the first image as foreground or background according to whether the sum of selected higher order DCT coefficients decreases or increases for the equivalent regions of the second image.

30   In the present context respective regions of two images of nominally the same scene are said to be equivalent if, in the case where the two images have the same resolution, the two regions correspond to substantially the same part of the scene or if, in the case where one image has a greater resolution than the other image, the part of the scene corresponding to the

3

region of the higher resolution image is substantially wholly contained within the part of the scene corresponding to the region of the lower resolution image.

If the two images are not substantially identical, due, for example, to slight camera movement, an additional stage of aligning the two images may be required.

6

Preferably, where the first and second images are captured by a digital camera, the first image is a relatively high resolution image, and the second image is a relatively low resolution pre- or post-view version of the first image.

When the image is captured by a digital camera, the processing may be done in the camera as

12   a post processing stage, i.e. after the main image has been stored, or as a post processing stage externally in a separate device such as a personal computer or a server computer. In the former case, the two DCT-coded images can be stored in volatile memory in the camera only for as long as they are needed for foreground/background segmentation and final image production. In the latter case, however, both images are preferably stored in non-volatile memory. In the case where a lower resolution pre- or post-view image is used, the lower

18   resolution image may be stored as part of the file header of the higher resolution image.

In some cases only selected regions of the two images need to be compared. For example, if it is known that the images contain a face, as determined, for example, by a face detection algorithm, the present technique can be used just on the region including and surrounding the face to increase the accuracy of delimiting the face from the background.

24

The present invention uses the inherent frequency information which DCT blocks provide and takes the sum of higher order DCT coefficients for a DCT block as an indicator of whether a block is in focus or not. Blocks whose higher order frequency coefficients drop when the main subject moves out of focus are taken to be foreground with the remaining blocks representing background or border areas. Since the image acquisition and storage

30   process in a conventional digital camera codes the captured images in DCT format as an intermediate step of the process, the present invention can be implemented in such cameras without substantial additional processing.

4

This technique is useful in cases where the differentiation created by camera flash, as described in PCT Application No. PCT/EP2006/005109, may not be sufficient. The two techniques may also be advantageously combined to supplement one another.

6   The method of the invention lends itself to efficient in-camera implementation due to the relatively simple nature of calculations needed to perform the task.

In a second aspect of the invention, there is provided a method of determining an orientation of an image relative to a digital image acquisition device based on a foreground/background analysis of two or more images of a scene.

12   BRIEF DESCRIPTION OF DRAWINGS

Embodiments of the invention will now be described, by way of example, with reference to the accompanying drawings, in which:

FIG. 1 is a block diagram of a camera apparatus operating in accordance with embodiments of the present invention.

18   FIG. 2 shows the workflow of a method according to an embodiment of the invention.

FIG. 3 shows a foreground/background map for a portrait image.

DESCRIPTION OF PREFERRED EMBODIMENTS

FIG. 1 shows a block diagram of an image acquisition device 20 operating in accordance with

24   embodiments of the present invention. The digital acquisition device 20, which in the present embodiment is a portable digital camera, includes a processor 120. It can be appreciated that many of the processes implemented in the digital camera may be implemented in or controlled by software operating in a microprocessor, central processing unit, controller, digital signal processor and/or an application specific integrated circuit, collectively depicted as block 120 labelled "processor". Generically, all user interface and control of peripheral

30   components such as buttons and display is controlled by a microcontroller 122. The processor 120, in response to a user input at 122, such as half pressing a shutter button (pre-capture mode 32), initiates and controls the digital photographic process. Ambient light exposure is determined using a light sensor 40 in order to automatically determine if a flash is

to be used. The distance to the subject is determined using a focusing mechanism 50 which also focuses the image on an image capture device 60. If a flash is to be used, processor 120 causes a flash device 70 to generate a photographic flash in substantial coincidence with the recording of the image by the image capture device 60 upon full depression of the shutter button. The image capture device 60 digitally records the image in colour. The image

6   capture device is known to those familiar with the art and may include a CCD (charge coupled device) or CMOS to facilitate digital recording. The flash may be selectively generated either in response to the light sensor 40 or a manual input 72 from the user of the camera. The high resolution image recorded by image capture device 60 is stored in an image store 80 which may comprise computer memory such a dynamic random access memory or a non-volatile memory. The camera is equipped with a display 100, such as an

12 LCD, for preview and post-view of images.

In the case of preview images which are generated in the pre-capture mode 32 with the shutter button half-pressed, the display 100 can assist the user in composing the image, as well as being used to determine focusing and exposure. Temporary storage 82 is used to store one or plurality of the preview images and can be part of the image store 80 or a

18 separate component. The preview image is usually generated by the image capture device 60. For speed and memory efficiency reasons, preview images usually have a lower pixel resolution than the main image taken when the shutter button is fully depressed, and are generated by subsampling a raw captured image using software 124 which can be part of the general processor 120 or dedicated hardware or combination thereof. Depending on the settings of this hardware subsystem, the pre-acquisition image processing may satisfy some

24 predetermined test criteria prior to storing a preview image. Such test criteria may be chronological, such as to constantly replace the previous saved preview image with a new captured preview image every 0.5 seconds during the pre-capture mode 32, until the final high resolution image is captured by full depression of the shutter button. More sophisticated criteria may involve analysis of the of the preview image content, for example, testing the image for changes, before deciding whether the new preview image should replace a

30 previously saved image. Other criteria may be based on image analysis such as the sharpness, or metadata analysis such as the exposure condition, whether a flash will be used for the final image, the estimated distance to the subject, etc.

6

If test criteria are not met, the camera continues by capturing the next preview image while discarding preceding captured preview image. The process continues until the final high resolution image is acquired and saved by fully depressing the shutter button.

Where multiple preview images can be saved, a new preview image will be placed on a chronological First In First Out (FIFO) stack, until the user takes the final picture. The reason for storing multiple preview images is that the last preview image, or any single preview image, may not be the best reference image for comparison with the final high resolution image in, for example, a red-eye correction process or, in the present embodiments, portrait mode processing. By storing multiple images, a better reference image can be achieved, and a closer alignment between the preview and the final captured image can be achieved in an alignment stage discussed later.

The camera is also able to capture and store in the temporary storage 82 one or more low resolution post-view images when the camera is in portrait mode, as will be to be described. Post-view images are essentially the same as preview images, except that they occur after the main high resolution image is captured.

In this embodiment the camera 20 has a user-selectable mode 30. The user mode 30 is one which requires foreground/background segmentation of an image as part of a larger process, e.g. for applying special effects filters to the image or for modifying or correcting an image. Thus in the user mode 30 the foreground/background segmentation is not an end in itself; however, only the segmentation aspects of the mode 30 are relevant to the invention and accordingly only those aspects are described herein.

If user mode 30 is selected, when the shutter button is depressed the camera is caused to automatically capture and store a series of images at close intervals so that the images are nominally of the same scene. The particular number, resolution and sequence of images, and the extent to which different parts of the image are in or out of focus, depends upon the particular embodiment, as will be described. A user mode processor 90 analyzes and processes the stored images according to a workflow to be described. The processor 90 can be integral to the camera 20 - indeed, it could be the processor 120 with suitable programming - or part of an external processing device 10 such as a desktop computer. In

this embodiment the processor 90 processes the captured images in DCT format. As
explained above, the image acquisition and storage process in a conventional digital camera
codes and temporarily stored the captured images in DCT format as an intermediate step of
the process, the images being finally stored in, for example, jpg format. Therefore, the
intermediate DCT-coded images can be readily made available to the processor 90.

FIG. 2 illustrates the workflow of an embodiment of user mode processing according to the
invention.

First, user mode 30 is selected, step 200. Now, when the shutter button is fully depressed, the
camera automatically captures and stores two digital images in DCT format:

- a high pixel resolution image (image A), step 202. This image has a foreground subject
  of interest which is in focus, or at least substantially more in focus than the
  background.
- a low pixel resolution post-view image (image B), step 204. This image has its
  background in focus, or at least substantially more in focus than the foreground
  subject of interest. Auto-focus algorithms in a digital camera will typically provide
  support for off-centre multi-point focus which can be used to obtain a good focus on
  the background. Where such support is not available, the camera can be focussed at
  infinity.

These two images are taken in rapid succession so that the scene captured by each image is
nominally the same.

In this embodiment steps 200 to 206 just described necessarily take place in the camera 20.
The remaining steps now to be described can take place in the camera or in an external device
10.

Images A and B are aligned in step 206, to compensate for any slight movement in the
subject or camera between taking these images. Alignment algorithms are well known.
Then, step 208, a high frequency (HF) map of the foreground focussed image A is
constructed by taking the sum of selected higher order DCT coefficients for each, or at least

8

the majority of, the DCT blocks of the image. By way of background, for an 8x8 block of pixels, a set of 64 DCT coefficients going from the first (d.c.) component to the highest frequency component is generated. In this embodiment, the top 25% of the DCT coefficients for a block are added to provide an overall HF index for the block. If not all the DCT blocks of the image are used to construct the map, those that are should be concentrated on the

6   regions expected to contain the foreground subject of interest. For example, the extreme edges of the image can often be omitted, since they will almost always be background. The resultant map is referred to herein as Map A.

Next, step 210, an HF map (Map B) of the background focussed image B is constructed by calculating the HF indices of the DCT blocks using the same procedure as for Map A.

12

Now, step 212, a difference map is constructed by subtracting Map A from Map B. This is done by subtracting the HF indices obtained in step 208 individually from the HF indices obtained in step 210. Since Image A has a higher pixel resolution than image B, a DCT block in Image B will correspond to a larger area of the scene than a DCT block in Image A. Therefore, each HF index of Map A is subtracted from that HF index of Map B whose DCT

18   block corresponds to an area of the scene containing or, allowing for any slight movement in the subject or camera between taking the images, substantially containing the area of the scene corresponding to the DCT block of Map A. This means that the HF indices for several adjacent DCT blocks in Image A will be subtracted from the same HF index of Map B, corresponding to a single DCT block in Image B.

24   At step 214, using the values in the difference map, a digital foreground/background map is constructed wherein each DCT block of Image A is assigned as corresponding to a foreground or background region of the image according to whether the difference between its HF index and the HF index of the DCT block of Image B from which it was subtracted in step 212 is respectively negative or positive.

30   Finally, step 216, additional morphological, region filling and related image processing techniques, alone or combination with other foreground/background segmentation techniques, can further improve and enhance the final foreground/background map.

The final foreground/background map 218 may now be applied to the DCT-coded or jpg version of Image A for use in processing the image according to the function to be performed by the user-selectable mode 30.

Where the processor 90 is integral to the camera 20, the final processed jpg image may be displayed on image display 100, saved on a persistent storage 112 which can be internal or a removable storage such as CF card, SD card or the like, or downloaded to another device, such as a personal computer, server or printer via image output device 110 which can be tethered or wireless. In embodiments where the processor 90 is implemented in an external device 10, such as a desktop computer, the final processed image may be returned to the camera 20 for storage and display as described above, or stored and displayed externally of the camera.

Variations of the foregoing embodiment are possible. For example, Image B could be a low resolution preview image rather than a post-view image. Alternatively, both Images A and B could be high resolution images having the same resolution. In that case a DCT block in Image B will correspond to the same area of the scene as a DCT block in Image A. Thus, in step 212, the difference map would be constructed by subtracting each HF index of Map A from a respective different HF index of Map B, i.e. that HF index of Map B corresponding to the same or, allowing for any slight movement in the subject or camera between taking the images, substantially the same area of the scene. In another embodiment both Images A and B are low resolution preview and/or post-view images having the same resolution, and the foreground/background map derived therefrom is applied to a third, higher resolution image of nominally the same scene. In a still further embodiment Images A and B have different pixel resolutions, and prior to DCT coding the pixel resolutions of the two images are matched by up-sampling the image of lower resolution and/or sub-sampling the image of higher resolution.

Although the embodiment described above contemplates the creation and storage of a digital foreground/background map, it may be possible to use the foreground/background designation of the image region corresponding to each DCT block directly in another algorithm, so that the formal creation and storage of a digital map is not necessary.

In another embodiment, rather than basing the maps and comparison on a DCT block by block analysis, each map can first be pre-processed to provide regions, each having similar HF characteristics. For example, contiguous blocks with HF components above a given threshold are grouped together and contiguous blocks with HF components below a given threshold are grouped together. Regions from the foreground and background images can then be compared to determine if they are foreground or background.

As mentioned above, the ability to provide foreground/background separation in an image is useful in many applications.

In a further aspect of the present invention, a particular application using a foreground/background map of an image, regardless of whether it has been calculated using the embodiment described above or for example using the flash-based technique of PCT/EP2006/005109, is to detect the orientation of an image relative to the camera. (The technique is of course applicable to any digital image acquisition device.) For most situations, this also implies the orientation of the camera when the image was taken without the need for an additional mechanical device.

Referring to Fig. 3, this aspect of the invention is based on the observation that in a normally oriented camera for a normally oriented scene, the close image foreground (in this case the subject 30) is at the bottom of the image and the far background is at its top.

Using flash-based foreground/background segmentation, being closer to the camera, the close foreground 30 reflects the flash more than the far background. Thus, by computing the difference between a flash and non-flash version image of the scene, the image orientation can be detected and camera orientation implied. (A corresponding analysis applies when analysing the DCT coefficients of two images as in the above described embodiment.)

An exemplary implementation uses 2 reference images (or preview images or combination of previous and main image suitably matched in resolution), one flash and one non-flash and transforms these into grey level.

For each pixel, the grey level of the non-flash image is subtracted from the one corresponding
6   to the flash image to provide a difference image. In other implementations, a ratio could be used instead of subtraction.

For each potential image/camera orientation direction, a box is taken in the difference image. So for an image sensing array 10 in an upright camera, box 12 is associated with an upright orientation of the camera, box 16 with an inverted orientation of the camera, box 14 with a
12   clockwise rotation of the camera relative to a scene and box 18 with an anti-clockwise rotation of the camera relative to the scene.

For each box 12-18, an average value of the difference image is computed. As such, it will be seen that in some implementations, the difference need only be calculated for portions of the image corresponding to the boxes 12-18.
18

For clarity, the boxes of Fig. 3 are not shown as extending to the edges of the image, however, in an exemplary implementation, for a box size = dim, the box 18 would extend from: left = 0, top = 0 to right = dim and bottom = image height. In other implementations, one could associate other suitable regions with a given orientation or indeed other units of measurement instead of the average (i.e. histograms).
24

The maximum of the average values for the boxes 12-18 is computed and the box corresponding to the largest value is deemed to be a region with the greatest degree of foreground vis-a-vis the remaining regions. This is deemed to indicate that this region lies at the bottom of the reference image(s). In the example of Fig. 3, the largest difference in the difference images of the boxes should occur in box 12, so indicating an upright subject and

12

implying an upright camera orientation given the normal pose of a subject. In some implementations the box 16 need not be used as it is not a realistic in-camera orientation.

In some implementations it can be of benefit to run some tests in order to validate the presumptive image orientation. For example, the maximum of the average values is tested to determine if is dominant vis-à-vis the other values and a level of confidence can be implied from this dominance or otherwise. The degree of dominance required can be varied experimentally for different types of images (indoor/outdoor as described in PCT/EP2006/005109, day/night). Information from other image analysis components which are used within the camera may be combined in this step for determining level of confidence. One exemplary image analysis component is a face tracking module which is operable on a stream of preview images. This component stores historical data relating to tracked face regions, including a confidence level that a region is a face and an associated orientation. Where multiple faces are present their data may be combined in determining a level of confidence.

If the difference values for the presumed left and right sides of an image are similar and smaller then the presumed bottom and larger than the presumed top, then it is more likely that the orientation has been detected correctly.

Because foreground/background maps can be provided for both indoor and outdoor images according to whether the maps have been created using flash or non-flash based segmentation, knowing image orientation can be useful in many further camera applications. For example, knowing the likely orientation of objects in an image reduces the processing overhead of attempting to identify such objects in every possible orientation.

The invention is not limited to the embodiments described herein which may be modified or varied without departing from the scope of the invention.

Claims

1.   A method of distinguishing between foreground and background regions of a digital image of a scene, the method comprising capturing first and second images of nominally the same scene and storing the captured images in DCT-coded format, the first image being taken with the foreground more in focus than the background and the second image being taken with the background more in focus than the foreground, and assigning regions of the first image as foreground or background according to whether the sum of selected higher order DCT coefficients decreases or increases for the equivalent regions of the second image.

2.   A method of distinguishing between foreground and background regions of a digital image of a scene, the method comprising the following steps:

(a) capturing first and second images of nominally the same scene and storing the captured images in DCT-coded format, the first image being taken with the foreground more in focus than the background and the second image being taken with the background more in focus than the foreground,

(b) calculating the sum of selected higher order coefficients for a plurality of DCT blocks of the first image,

(c) calculating the sum of the same higher order coefficients for a plurality of DCT blocks of the second image,

(d) comparing the sum calculated in step (b) for each DCT block of the first image with the sum calculated in step (c) for the DCT block of the equivalent region of the second image,

(e) if the sum calculated for a given DCT block of the first image is greater than the sum calculated for the DCT block of the equivalent region of the second image, assigning the given block as corresponding to a foreground region of the image, and

(f)   if the sum calculated for a given DCT block of the first image is less than the sum calculated for the DCT block of the equivalent region of the second image, assigning that block as corresponding to a background region of the image.

3.   The method of claim 2 further comprising the step of aligning the first and second images between step (a) and step (b).

14

4.   The method of claim 2 further comprising creating a digital map from the result of assigning said blocks as corresponding to a background or foreground region of the image.

5.   The method of claim 2, wherein the first and second images have different pixel resolutions.

6.   The method of claim 5, further comprising matching the pixel resolutions of the two images prior to DCT coding.

7.   The method of claim 2, wherein the first and second images are captured by a digital camera.

8.   The method of claim 7, wherein the first image is a relatively high resolution image, and wherein the second image is a relatively low resolution pre- or post-view version of the first image.

9.   The method of claim 7, wherein steps (b) to (f) are performed as a post processing stage in a device external to the digital camera.

10.   The method of claim 7, wherein steps (b) to (f) are performed as a post processing stage in the digital camera.

11.   The method of claim 2, wherein the first and second images have the same pixel resolution.

12.   The method of claim 11, wherein the first and second images are captured by a digital camera and are relatively low resolution pre- and/or post-view versions of a higher resolution image of said scene also captured by the camera.
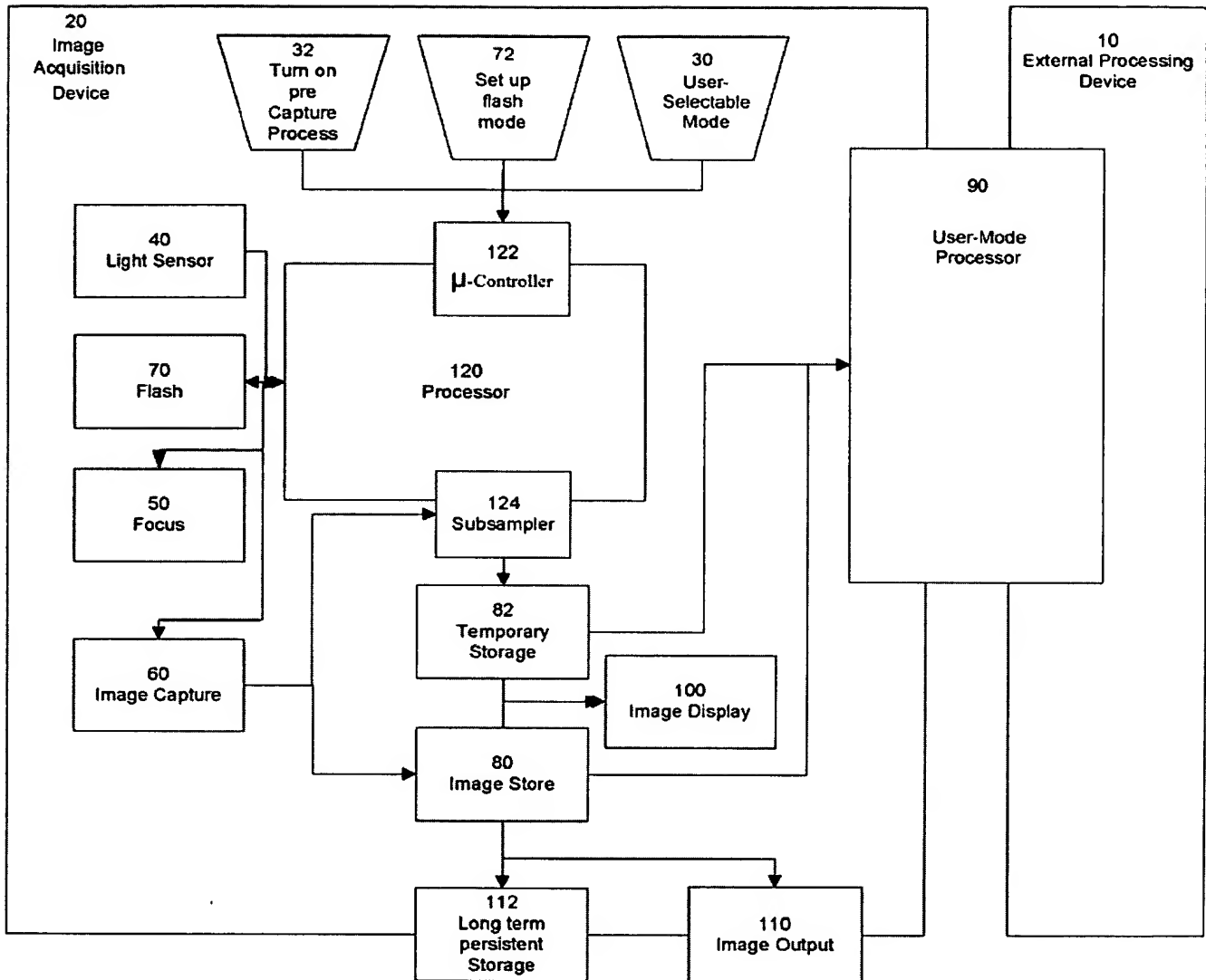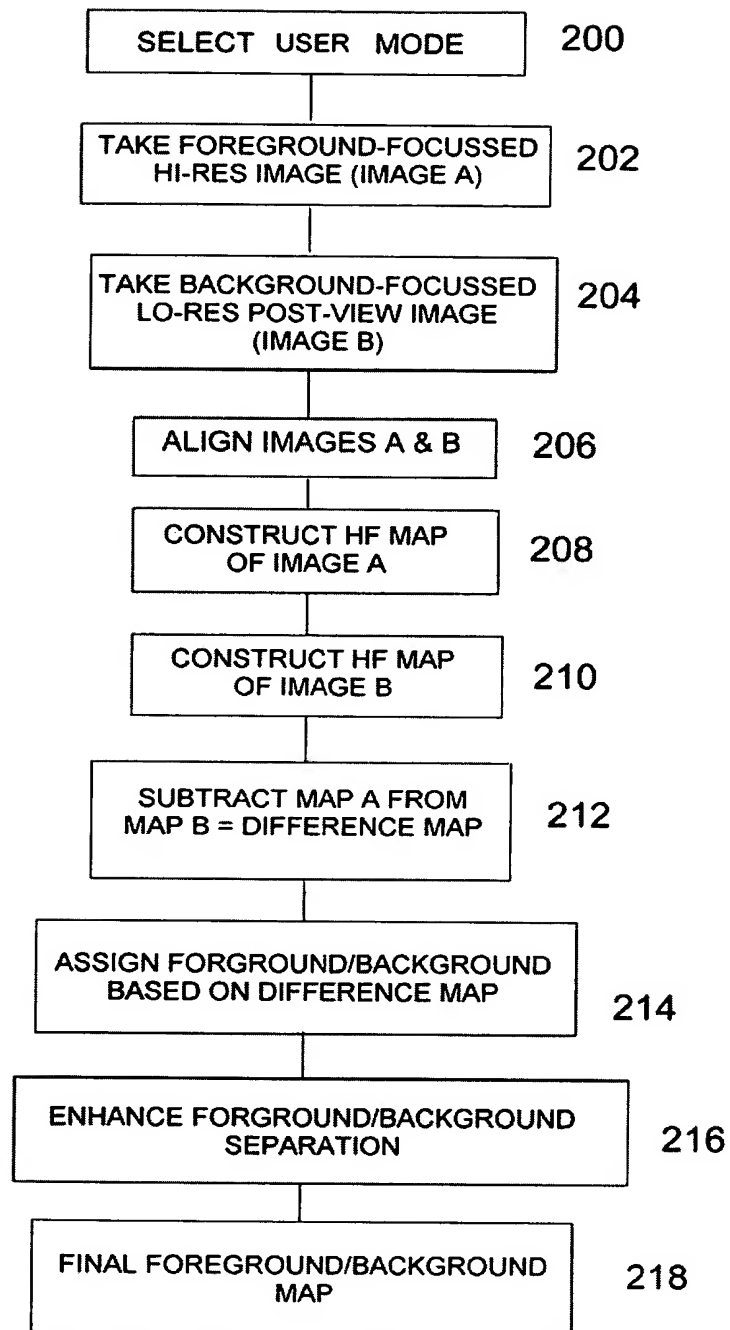
13.   The method claimed in claim 2, wherein the selected higher order coefficients are the top 25% of coefficients.

14.   A digital image acquisition system having no photographic film comprising means for capturing first and second images of nominally the same scene and storing the captured images in DCT-coded format, the first image being taken with the foreground more in focus than the background and the second image being taken with the background more in focus than the foreground, and means for assigning regions of the first image as foreground or background according to whether the sum of selected higher order DCT coefficients decreases or increases for the equivalent regions of the second image.

15.   A method of determining an orientation of an image relative to a digital image acquisition device, comprising:

capturing two images nominally of the same scene with said digital image acquisition device;

comparing at least a portion of said two images adjacent the corresponding edges of said images to determine whether said portion comprises relatively more foreground than background; and

responsive to said portion comprising more than a threshold degree of foreground, determining that said images are oriented with said portion at their bottom.

16.   A method as claimed in claim 15 in which said images comprise a flash image and a non-flash image and in which said comparing comprises comparing the luminance levels of the pixels of said portion.

17.   A method as claimed in claim 15 in which said images comprise non-flash images and in which said comparing comprises comparing higher order DCT coefficients for at least one block of said portion.

18.   A method as claimed in claim 15 comprising implying an orientation of said digital image acquisition device in accordance with said image orientation.

19.   A method as claimed in claim 15 wherein said comparing comprises comparing respective portions adjacent a plurality of edges of said two images, and wherein a portion

16

which is determined to include the greatest degree of foreground relative to any other portions is deemed to be located at the bottom of said images.

20.   A method as claimed in claim 19 wherein for a portion to be deemed to be located at the bottom of said images, its degree of foreground must exceed the degree of foreground for a portion adjacent an opposite edge by a given threshold.

21.   A method as claimed in claim 20 wherein said threshold is varied according to at least one of: the exposure level of said images and whether said images are classified as being indoor or outdoor.

22.   A method as claimed in claim 19 wherein for portion to be deemed to be located at the bottom of said images, its degree of foreground must exceed the degree of foreground for a portion adjacent at least an adjacent edge, and the degree of foreground for the portion adjacent said adjacent edge must exceed the degree of foreground for a portion adjacent an opposite edge.

23.   A digital image acquisition system having no photographic film comprising: means for capturing two images nominally of the same scene; means for comparing at least a portion of said two images adjacent the corresponding edges of said images to determine whether said portion comprises relatively more foreground than background; and means, responsive to said portion comprising more than a threshold degree of foreground, for determining that said images are oriented with said portion at their bottom.
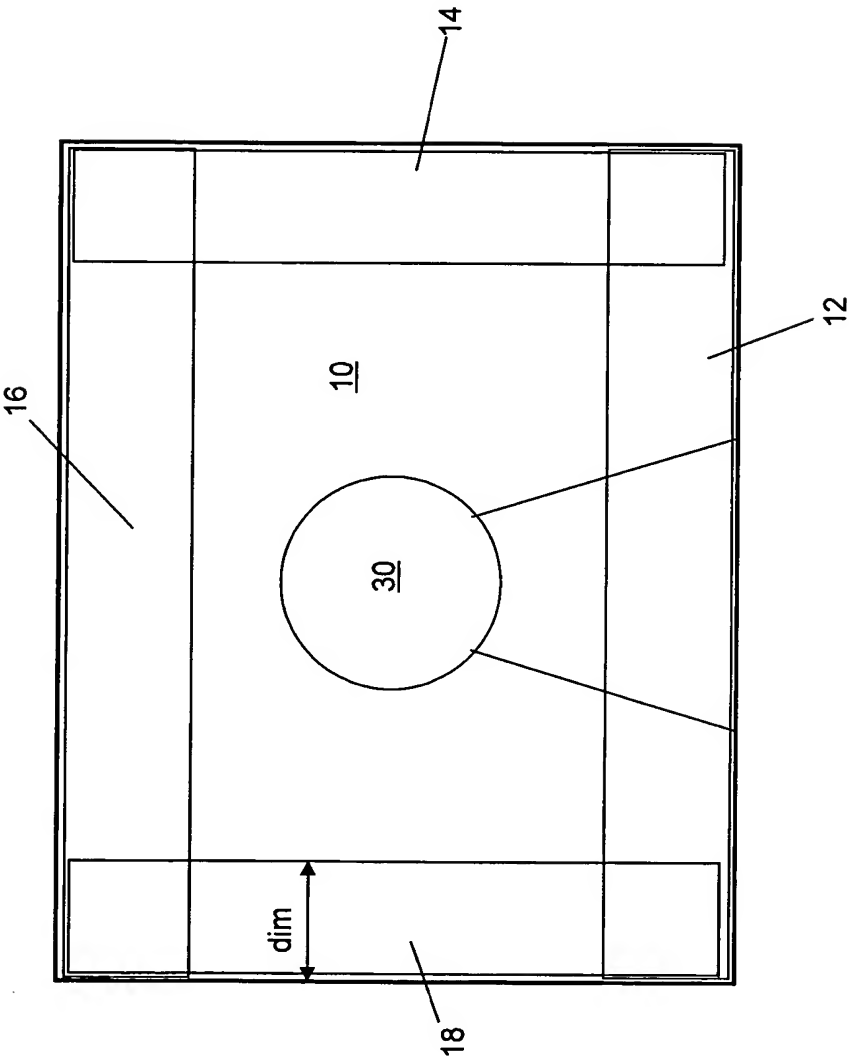
Figure 1

```
┌─────────────────────────────┐
│     SELECT  USER  MODE       │   200
└─────────────────────────────┘
              │
┌─────────────────────────────┐
│  TAKE FOREGROUND-FOCUSSED    │   202
│  HI-RES IMAGE (IMAGE A)      │
└─────────────────────────────┘
              │
┌─────────────────────────────┐
│  TAKE BACKGROUND-FOCUSSED    │   204
│  LO-RES POST-VIEW IMAGE      │
│        (IMAGE B)             │
└─────────────────────────────┘
              │
┌─────────────────────────────┐
│     ALIGN IMAGES A & B       │   206
└─────────────────────────────┘
              │
┌─────────────────────────────┐
│     CONSTRUCT HF MAP         │   208
│       OF IMAGE A             │
└─────────────────────────────┘
              │
┌─────────────────────────────┐
│     CONSTRUCT HF MAP         │   210
│       OF IMAGE B             │
└─────────────────────────────┘
              │
┌─────────────────────────────┐
│    SUBTRACT MAP A FROM       │   212
│  MAP B = DIFFERENCE MAP      │
└─────────────────────────────┘
              │
┌─────────────────────────────┐
│ ASSIGN FORGROUND/BACKGROUND  │
│  BASED ON DIFFERENCE MAP     │   214
└─────────────────────────────┘
              │
┌─────────────────────────────┐
│ ENHANCE FORGROUND/BACKGROUND │   216
│        SEPARATION            │
└─────────────────────────────┘
              │
┌─────────────────────────────┐
│ FINAL FOREGROUND/BACKGROUND  │   218
│           MAP                │
└─────────────────────────────┘
```

Figure 2

FIG. 3